Overview of the NLPCC 2018 Shared Task: Social Media User Modeling

Fuzheng Zhang[1(✉)] and Xing Xie[2]

[1] Meituan AI Lab, Beijing, China
zhangfuzheng@meituan.com
[2] Microsoft Research Asia, Beijing, China
xingx@microsoft.com

Abstract. In this paper, we give the overview of the social media user modeling shared task in the NLPCC 2018. We first review the background of social media user modeling, and then describe two social media user modeling tasks in this year's NLPCC, including the construction of the benchmark datasets and the evaluation metrics. The evaluation results of submissions from participating teams are presented in the experimental part.

Keywords: User modeling · Social media · Recommendation

1 Background

With the widespread of social media websites in the internet, and the huge number of users participating and generating infinite number of contents in these websites, the need for personalization increases dramatically to become a necessity. One of the major issues in personalization is building users' profiles, which depend on many elements; such as the used data [1, 2], the application domain they aim to serve [3, 4], the representation method and the construction methodology [5, 6]. Another major issue in personalization is personalized recommendation, which can be divided into different methods including contented based methods [7, 8], collaborative filtering based methods [9, 10], and hybrid methods [11, 12].

In the industry field, many influential user modeling products have been built, such as Netflix movie recommendation system, Amazon item recommendation system, etc. These kinds of systems are immerging into every user's life.

Under such circumstance, in this year's NLPCC shared task, we call the social media user modeling task that cover both personalized recommendation and user profiling tasks.

The remainder of this paper is organized as follows. Section 2 describes the provided dataset. In Sect. 3, we describe the detail of these two shared tasks. Section 4 describes evaluation metrics, and Sect. 5 presents the evaluation results of different submissions. We conclude the paper in Sect. 6, and point out our plan on future user modeling activities.

M. Zhang et al. (Eds.): NLPCC 2018, LNAI 11109, pp. 464–467, 2018.
https://doi.org/10.1007/978-3-319-99501-4_45

2 Data Description

The data, collected from a social media platform, contains the following five aspects:

(1) profile.txt describes users' profiles. Currently gender, province and city are provided.

user	gender	province	city	tags

(2) tags.txt describes users' tags. Each line contains a user and related tag.

user	tag

(3) social.txt describes users' following relationship, where user1 follows user2 on this social media platform.

user 1	user2

(4) tweets.txt describes what user posted. Each line contains a user and the posted tweet.

user	tweet

(5) checkins.txt describes users' location visits. The format is as follows, where POI is the location user visits, cate1, cate2, cate3 is the category of the POI in a hierarchical level. lat and lng is the latitude and longitude information and Name is the location name.

user	POI	cate1	cate2	cate3	lat	lng	name

All the files are UTF-8 encodes and tab separated.

3 Task Description

Given the social media dataset including the following heterogenous information: users' profiles (gender, province, city, tags), social ties (following relationship), users' published tweets, and users' location visits, the NLPCC 2018 social media user modeling shared task includes two shared tasks for social media dataset: User Tags Prediction task and User Following Recommendation task.

3.1 User Tags Prediction Task

Given users' other information except tags, predict which tags are related to a user.

3.2 User Following Recommendation Task

Given users' following relationship and other provided information, predict the users a user would like to follow in the future.

4 Evaluation Metrics

The quality of **User Tags Prediction (UTP)** and **User Following Recommendation (UFR)** subtasks will both be evaluated by F1@K,

$$\boldsymbol{P_i@K} = \frac{|\boldsymbol{H_i}|}{\boldsymbol{K}}, \quad \boldsymbol{R_i@K} = \frac{|\boldsymbol{H_i}|}{|\boldsymbol{V_i}|}, \quad \mathbf{F1}_{\boldsymbol{i}}\boldsymbol{@K} = \frac{\boldsymbol{P_i@K} * \boldsymbol{R_i@K}}{\boldsymbol{P_i@K} + \boldsymbol{R_i@K}}$$

$$\mathbf{F1}\boldsymbol{@K} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \mathbf{F1}_{\boldsymbol{i}}\boldsymbol{@K}$$

where $|H_i|$ is the correctly predicted item set (item refers to tag in UTP and user in UFR) for user i 's top K prediction, $|V_i|$ is the ground truth item set for user i. $P_i@K$, $R_i@K$ and $F1_i@K$ is the precision, recall and F1 for a user i.

In **UTP**, we set $K = 3$.

In **UFR**, we set $K = 10$.

5 Evaluation Results

There are totally 39 teams registered for the above two shared tasks, and 4 teams submitted their results for UTP subtask and 3 teams submitted their results for UFR subtask. Table 1 and Table 2 lists the evaluation results respectively.

Table 1. Evaluation results of the User Tags Prediction (UTP) subtask.

	F1@3
Team 1	0.046268361
Team 2	0.033795288
Team 3	0.000124505
Team 4	0

Table 2. Evaluation results of the User Following Recommendation (UFR) subtask.

	Accuracy
Team 1	0.011645037
Team 2	0.004177592
Team 3	0.00000101

6 Conclusion

This paper briefly introduces the overview of this year's two social media user modeling shared tasks. We have 39 teams registered and 4 teams submitted final submissions. In the future, we plan to provide more social media datasets and call for new user modeling tasks.

References

1. Zhong, Y., Yuan, N.J., Zhong, W., Zhang, F., Xie, X.: You are where you go: inferring demographic attributes from location check-ins. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, pp. 295–304 (2015)
2. Abel, F., Gao, Q., Houben, G.-J., Tao, K.: Semantic enrichment of twitter posts for user profile construction on the social web. In: Extended Semantic Web Conference, pp. 375–389 (2011)
3. Zhang, F., Zheng, K., Yuan, N.J., Xie, X., Chen, E., Zhou, X.: A novelty-seeking based dining recommender system. In: The International Conference, pp. 1362–1372 (2015)
4. Morita, M., Shinoda, Y.: Information filtering based on user behavior analysis and best match text retrieval. In: Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 272–281 (1994)
5. Adomavicius, G., Tuzhilin, A.: User profiling in personalization applications through rule discovery and validation. In: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 377–381 (1999)
6. Zhang, F., Yuan, N.J., Lian, D., Xie, X.: Mining novelty-seeking trait across heterogeneous domains. In: Proceedings of the 23rd International Conference on World Wide Web, pp. 373–384 (2014)
7. Pazzani, M.J., Billsus, D.: Content-based recommendation systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72079-9_10
8. Popescul, A., Pennock, D.M., Lawrence, S.: Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In: Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, pp. 437–444 (2001)
9. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Comput. (Long. Beach. Calif.) **42**(8), 30–37 (2009)
10. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: Bayesian personalized ranking from implicit feedback. In: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, pp. 452–461 (2009)
11. Zhang, F., Yuan, N.J., Lian, D., Xie, X., Ma, W.-Y.: Collaborative knowledge base embedding for recommender systems. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 353–362 (2016)
12. Zhang, F., Yuan, N.J., Zheng, K., Lian, D., Xie, X., Rui, Y.: Exploiting dining preference for restaurant recommendation. In: Proceedings of the 25th International Conference on World Wide Web, pp. 725–735 (2016)